

2022

Artificial Knowing Otherwise

Os Keyes

University of Washington

okeyes@uw.edu

Kathleen A. Creel

Northeastern University

k.creel@northeastern.edu

Recommended Citation

Keyes, Os, and Kathleen A. Creel. 2022. "Artificial Knowing Otherwise." *Feminist Philosophy Quarterly* 8 (3/4). Article 12.

Artificial Knowing Otherwise

Os Keyes and Kathleen A. Creel¹

Abstract

While feminist critiques of AI are increasingly common in the scholarly literature, they are by no means new. Alison Adam's *Artificial Knowing* (1998) brought a feminist social and epistemological stance to the analysis of AI, critiquing the symbolic AI systems of her day and proposing constructive alternatives. In this paper, we seek to revisit and renew Adam's arguments and methodology, exploring their resonances with current feminist concerns and their relevance to contemporary machine learning. Like Adam, we ask how new AI methods could be adapted for feminist purposes and what role new technologies might play in addressing concerns raised by feminist epistemologists and theorists about algorithmic systems. In particular, we highlight distributed and federated learning as providing partial solutions to the power-oriented concerns that have stymied efforts to make machine learning systems more representative and pluralist.

Keywords: feminist epistemology, machine learning, sociotechnical systems, distributed learning, federated learning, Alison Adam, artificial intelligence

1. Introduction

In the early 1980s, renewed optimism about artificial intelligence fueled efforts to expand the reasoning abilities of artificial systems beyond logic to include common sense. At the core of one such AI system, named Cyc, was a vast knowledge base built fact by fact by human "ontologists" or knowledge engineers. The engineers encoded propositions ranging from the relationship between meters and kilometers to "common knowledge" such as "you are not likely to get a speeding ticket in mid- or late-twentieth century America if you're driving less than 5 m.p.h. over the speed limit" (Adam 1998, 88; citing Lenat and Guha 1990, 284). The latter claim was internally tagged as "knowledge" rather than "belief," a designation reflecting the designers' contention that it was uncontroversial (Adam 1998, 88).

As feminist philosopher Alison Adam argued at the time, this and many other similarly tagged facts only appear to be uncontroversial from the perspective of the

¹ Equal contributors.

builders. Cyc's knowledge base purported to represent a universal perspective, a "view from nowhere" (Nagel 1989), while in fact it presented the perspectives of the predominantly white, middle class, male mathematicians who built it (Adam 1998; Code 2012). It hid the fact of its situatedness—the fact that its knowledge base represented, as its builders acknowledged, "TheWorldAsTheBuildersOfCycBelievelt-ToBe" (Adam 1998, 88). If gender, age, race, and type of car all affect how likely one is to receive a speeding ticket, mph over the limit is not enough to determine the likelihood of ticketing (Adam 1998, 89). Context affects the perceived truth of the claim. This claim may have been common knowledge, but it was not uncontroversial.

Claims of universality were common in "symbolic AI," the research program that sought to understand intelligence and human reason by building artificial systems to manipulate symbols using logic or procedural rules. In part due to critiques of symbolic AI's ontological model, most prominently by Hubert Dreyfus (1992), and the apparent validation of those critiques when the symbolic program stalled, the broader landscape of artificial intelligence has changed since the 1990s. Although the Cyc project itself remains active (Knight 2016), most contemporary AI does not rely on vast knowledge bases of propositional claims. Instead, machine learning distills statistical patterns from existing data without explicit guidance from human knowledge engineers.

Yet despite regular claims that "this time, it is different" (Wajcman 2017), to feminist scholars the political tendencies of conventional machine learning look familiar. Although data are no longer manually typed into knowledge bases by engineers, biases that favor socially dominant groups still creep into both data and models, as has been demonstrated by Noble (2018), Benjamin (2019), and others. Contemporary machine learning still makes claims to objectivity: it still is described as neutral, still inhabits the "view from nowhere" while in fact representing views of socially dominant groups. Although the particular "political orders" resulting from the incorporation of machine learning systems into political life are new (Amoore 2022), feminist concerns with the epistemology of machine learning are not.

Critiquing and redesigning algorithmic systems to better represent a plurality of views remains the work of feminist critics of artificial intelligence. Feminist scholars have examined the biased outcomes of algorithmic systems (Noble 2018; Hutson et al. 2018), highlighted the narrow range of knowers and ways of knowing involved in models (Keyes 2018; Stark 2018; Sadowski 2019), and drawn attention to the broader political economy in which widely used models are designed and deployed (Gray and Suri 2019; Bucher 2018). But feminist critiques of AI often focus on the disassembly of existing systems rather than the assembly of new ones. Even scholarship that highlights the need for creation, such as "data feminism" and the "Feminist Data

Manifest-no,” which involves a “commitment to new data futures” (Cifor et al. 2019), rarely focuses on imagining or building feminist algorithmic systems.²

Feminist scholarship that focuses on responding to prominent and consequential examples of algorithmic systems causing harm is essential.³ But to focus exclusively on existing systems causing harm would be to miss the opportunity to imagine what machine learning might be. We believe it is essential to engage in creative experimentation with alternatives to the models we critique, lest our map of critique tacitly become our sense of the territory.

Such positive engagements would ideally take the form of what Phillip E. Agre referred to as a “critical technical practice,” a way of working for change that involves “a split identity—one foot planted in the craft work of design and the other foot planted in the reflexive work of critique” (Agre 2014, 155). Such a practice would allow us to both reshape the epistemic frames in which algorithmic systems are developed, and practically test and implement feminist theories of epistemic justice and moral relation.

One feminist scholar of AI who has done just that is Alison Adam. Her work features not only adroit disassembly of the systems she was engaging with but also prototyping and exploration of “feminist AI projects.” Since her work twenty years ago, the technologies and ontological approaches underlying AI have shifted dramatically. But we believe that both her concerns and her recuperative impulse—her desire to not only critique but engage in “the more difficult task of thinking through the ways in which AI research could be informed by feminist theory” (Adam 1998, 156)—remain not only relevant but also under-explored subsequent to her work.

In this paper, we revisit Adam’s arguments and projects, exploring how they resonate with current feminist concerns about artificial intelligence methods. Like Adam, we also ask how new AI methods could be adapted for feminist purposes and what role newer technologies might play in ameliorating or addressing some of the concerns raised by feminist epistemologists and theorists about algorithmic systems. Focusing in particular on distributed and federated machine learning, we argue that they provide a partial solution to some of the power-oriented concerns that have stymied efforts to increase the representationality and plurality of machine learning systems’ underlying “knowers.”

² Exceptions include Hancox-Li and Kumar’s (2021) proposal for the use of feminist epistemology in algorithmic development and Lilly Irani’s “turkopticon” intervention (Irani and Silberman 2013).

³ In other work, we share the focus on critique of existing systems (Keyes 2020; Rincón, Keyes, and Cath 2022)

Adam hoped that her proposals might chart “a course between the Scylla of a ‘nothing changes’ pessimism and the Charybdis of a gushingly unrealistic ‘fabulous feminist future’ of artificial intelligence (Adam 1998, 156). In this paper, we aim to chart a similar course in proposing to repurpose current machine learning techniques to support an understanding of situated multiplicity.

2. Whose Knowledge?

2.1. Knowledge in Cyc and Soar

A central question in Alison Adam’s investigation of AI systems—one that has long motivated feminist epistemologists more generally—is that of whose knowledge is represented in systems’ models of the world. For Adam, as for many other feminist philosophers, this question is entangled with the question of what forms of knowledge are taken seriously. But aspects of the “who” can be disentangled from the “what”: the knowers taken seriously are clear even when the forms of knowledge require more analysis.

Adam’s inquiry into the “who” of AI’s knowledge focuses on two large symbolic AI projects, Cyc and Soar. Douglas Lenat founded Cyc in 1984 on the premise that building an enormous knowledge base of commonsense facts was the only way to train an intelligent machine. He hoped that Cyc’s performance would surpass then-common “expert systems,” symbolic AI systems combining a “knowledge base” of facts about medicine or law with rules-based inference engines that answered questions or added to the knowledge base. Previous expert systems built deep knowledge in their domains of focus. However, due to the narrowness of their expertise, they were prone to breaking unpredictably at the margins of their knowledge. For example, a medical expert system trained to recommend dosages of pain medication might lack “commonsense” knowledge about what might differentiate the injuries caused by falling off a roof from those resulting from a car crash. In order to overcome the “brittleness” of traditional expert systems, Cyc’s founders hoped to build a margin-less system (Adam 1998, 81). Due to the scope of its ambition, Cyc was first deemed ready to put into commercial use in 2016 (Knight 2016).

Soar, founded by Allen Newell, operates on the problem-solving model of its original acronym: State, Operator, and Result (Adam 1998, 91). Soar searches its problem states for a solution that matches its goal. Its strategies are based on careful study of human problem solvers. However, its only test subjects were young male undergraduate students at the Carnegie Institute of Technology (Adam 1998, 93).

In her investigations into both systems, Adam finds models of the ideal knower implicit in the design of the system. In the case of Cyc, the premise of the system is to articulate a “consensus reality, or the millions of things that we assume that everyone else knows” (Adam 1998, 83). As Adam emphasizes, this idea of a singular

“consensus” is impossible in practice; developers base their evaluation of what constitutes part of that reality on their own understandings of “the human” and of what propositions count as knowledge. It is those developers who are taken to have “an epistemologically authoritative ‘non-weird’ perspective on true knowledge of the world” (Adam 1998, 88), resulting in a situation where it is “middle-class, male, professional knowledge [that] informs TheWorldAsTheBuildersOfCycBelieveItToBe” (Adam 1998, 90).⁴

Soar, despite positioning itself as a response and alternative to Cyc, produces similar problems. Although it claims to be built on the basis of an empirically tested model of problem-solving and cognition, Adam notes that the developers treated the subjects of that empirical testing as irrelevant to the generalizability of their theory, raising the question of whose forms of thinking and problem-solving are seen as universal. While the resulting knowledge base does go further than simply extrapolating from the beliefs of the developers alone, the result is still a model of knowledge “based on the behaviour of a few technically educated, young, male, probably middle-class, probably white, college students working on a set of rather unnatural tasks in a US university in the late 1960s and early 1970s” (Adam 1998, 94).⁵

The question of knowledge ascription returns in Adam’s proposals for “feminist AI projects.” Its indirect presence can be found in her first example: a legal expert system designed to advise people on the law concerning an injury they have suffered. At that time, the most common model for expert legal systems was to evaluate whether a case would succeed. Taking into account both the structural misogyny of the legal system and the way this undercuts the self-trust and confidence of women encountering it, Adam’s system instead provides “examples of past cases which bear some resemblance to the present case [and so] leaves the question of whether or not to proceed open to the users, rather than making a decision for them” (Adam 1998, 160). While Adam’s concern here is primarily trust and agency, it is notable that her method of pursuing it implicitly brings the (woman) user into frame as a knower: as someone whose knowledge and judgment contributes to the evaluation of her case’s success.

More directly linked to representation is Adam’s second example, that of a natural language processing system explicitly modeled on what she sees as women’s forms of speech and “conversational repair.”⁶ As this suggests, her proposal is not

⁴ Sherron (2000) also critiques Cyc’s project of “constructing common sense” from a feminist perspective.

⁵ The fact that all but one of the participants was college-educated is just one illustration of the unrepresentative nature of the sample.

⁶ Adam is cognizant that there is no universal gendered form of speech and that “the model described here is a white, middle-class, Anglo-American English one, which

only explicitly cognizant of the cultural constraints it represents, but it is designed to incorporate knowledge and perspectives that fall outside conventional linguistic understandings of conversation that were predominantly premised on the practices of (white) men. In both cases, as different as they are, we find Adam seeking to develop systems that explicitly attend to the breadth of the knowledge they incorporate and the breadth of the claims they can make as a result.

2.2. “Whose Knowledge?” in the Present

The models of knowledge Adam analyses—monolithic ontologies of everything designed to underpin expert systems—may appear outmoded today, originating as they did in a different epistemology than today’s flexible and adaptive machine learning systems.⁷ But the issues she raises regarding whose knowledge underpins AI systems are, if anything, more pressing given the increasing prevalence of AI itself.

Researchers continue to highlight the narrow range of perspectives that the datasets underlying machine learning systems represent (Noble 2018; Keyes 2018). Machine learning’s reliance on free, large-scale resources (some prominent examples include Flickr content for facial and object recognition, Wikipedia for text analysis and image classification, and CommonCrawl for web pages) means that systems often represent only the knowledge and knowers recognized by existing infrastructures, each with their own partial cultural frame (Ford and Wajcman 2017). Further, even when trained with putatively “neutral” data, the problems AI is designed to address and the framings of those problems are deeply entangled with existing hierarchies of power (Mager 2014; Keyes 2020; Browne 2015; Stevens and Keyes 2021; Introna and Nissenbaum 2000). The result is ongoing disconnects between systems’ representation of the world as their developers “BelieveItToBe” and representations of the world as others believe it to be.

A concerned reader may be tempted to resolve such examples by ensuring greater representation of marginalized knowledge and knowers. Many machine learning practitioners have advocated just that. Notwithstanding questions of essentialism and stereotyping—of whether these efforts risk fixing in place “foundational” ideas of dynamic identities and lives—work focused on representation alone cannot fully address the broader, structural aspects of AI (Soon 2021).

We live not only in a world of increasing automation but also in a world where the terms of that automation and the choice of data underlying it are controlled by

probably does not even fit, for example, New York Jewish speech” (Adam 1998, 163), much less forms of speech any further afield from Adam’s own perspective.

⁷ In fact, monolithic ontologies are still common, as discussed in Vrandečić and Kröttsch (2014).

organizations that sit largely outside democratic mechanisms of accountability, control, and consent. Even absent biases in data, the cultural milieu in which software and AI development take place can produce and reinforce disparities (Allhutter 2019). Under such circumstances, calls for representation without other changes risk reinforcing these structures and approaches. Far from torpedoing the project of facial recognition, concerns about bias in facial recognition software have instead been recuperated by the technology companies developing these systems to justify folding further, more diverse populations into their surveillance network (Merler et al. 2019). Treating incorporation and representation as the only solution ignores the fact that there may be very good reasons to not make data available—not only in the case of surveillance systems but also in cases where continuing epistemic injustices make inclusion its own form of harm (Christen 2012).

These difficulties with representation are increasingly recognized, including by Catherine D’Ignazio and Lauren Klein. In their recent book *Data Feminism*, D’Ignazio and Klein (2020) warn against simplistic, representation-oriented “fixes” and describe projects that broaden the knowledge and knowers involved in datalogical thinking. Examples range from collaborative projects to map femicides to community-driven mapping programs. But none of these community-scale projects require machine learning to implement. Machine learning (ML) typically relies on big data, and gathering data of sufficient size can be challenging for small groups hoping to stage critical technical interventions using AI. The question becomes, then, whether there are plausible ways to build ML systems (such as Adam’s language project) that do not fall into the trap of transferring power to and endorsing the form of these wider structures.

2.3. Localization and Distribution: Critical Technical Practices

We believe there are plausible ways to build ML systems that do not fall into that trap, and that efforts to create such systems can build on recent developments in machine learning itself. Such efforts will also be founded on the premise that the issues to be addressed are *sociotechnical* and are best addressed with entangled technological and social approaches. Machine learning systems alone, while not agnostic, can be adapted to diverse purposes. This includes our own proposals, which should not be taken in isolation.

Our concern is the development of machine learning systems that learn from a more diverse range of knowers without concentrating data and power. In this section, we hope to offer a model that would allow for the representation of diverse knowers in a pluralistic machine learning system while simultaneously shielding those included from some of the risks of being data subjects incorporated into algorithmic systems. We propose an examination of multitask federated learning.

Many conventional forms of machine learning—found in both popular discourse and the practices of developers—imagine a centralizing algorithmic-development process. Data streams into a central hub where a single party develops and controls a single model. The model demands that data be “handed over” to a single, authoritative algorithmic interpreter to be analyzed on that interpreter’s terms. A return to Adam’s proposed language model, an effort to develop a system based on “feminine” forms of interpretation and conversational repair, suggests the concerns with this centralizing model. Using conventional machine learning, building Adam’s system based on “feminine” linguistic patterns would require first collecting and standardizing a vast array of examples of feminine speech, centralizing it, and using that centralized corpus to produce a single model capable of re-presenting the language patterns it has been exposed to. Such an approach raises concerns around accumulation of power, data extraction, and control that would heavily limit our willingness to call it feminist.

Likewise, many existing technical fixes for problems like violations of privacy improve on the status quo, decreasing the violation of privacy without significantly shifting the balance or distribution of power. Consider the case of privacy in large datasets, such as medical records, health information, or reviewer profiles. As Latanya Sweeney (2000) showed, 87 percent of the population of the United States in 1997 was uniquely identifiable in purportedly “anonymized” data that included records of zip codes, birthdays, and sex. Sweeney famously illustrated this by finding Massachusetts Governor William Weld’s data in the supposedly anonymized records of state employees released to health researchers by the Massachusetts Group Insurance Commission.⁸

Formal measures of privacy such as k -anonymity and differential privacy aim to solve this problem of reidentification. K -anonymity, for example, solves the aforementioned problem by ensuring that for each set of identifying features in the dataset, there are at least a certain number of people, identified with the variable k , who share those identifiers. For example, if there are at least three people in a standard anonymized medical dataset who share the same birthday and zip code, then $k = 3$ in that dataset.

Differential privacy solves the same problem by intentionally making minor modifications to the data, such as changing the day in a date of birth, in order to decrease the likelihood of uniquely identifying individuals in the data. In doing so, differential privacy “addresses concerns that any participant might have about the leakage of her personal information: even if the participant removed her data from

⁸ Such reidentifications are more difficult now due to the 2003 HIPAA Privacy Rule; see Barth-Jones (2012) for discussion.

the data set, no outputs . . . would become significantly more or less likely” (Dwork 2008, 2).

K-anonymity and differential privacy can protect individuals from accidental exposure when anonymized data is intentionally released. But as Philip Rogaway (2015) argues, these metrics imagine a world in which the threat to privacy comes exclusively from the person querying the database—the “adversary” interested in piecing together scraps of data to expose the privacy of individuals. The greatest threats to privacy may instead come from a source that formal metrics do not address: the compiling and indefinite maintenance of large databases that are perpetually at risk of being leaked in their entirety in a data breach,⁹ being queried by inside actors, or being surveilled by state agencies (Rogaway 2015, 20–21). Formal privacy measures like differential privacy do not measure the size of the dataset created, the number of people exposed if the data were to leak, or the concentration of access to the database. As such, formal privacy temporarily protects individual privacy without changing the fundamental risks and power imbalances of the system. While it may help some knowers to be represented in the system without exposing them to extractive data use, it does not otherwise change whose knowledge is represented or whose questions can be answered.

Consider, by contrast, a distributed learning paradigm. Rather than the standard centralized model of machine learning, in which data is collected so that a model can learn from it all together, distributed learning sends a naive machine learning model out into the world to learn from all the data it meets and to update its model on each stop of its digital journey. For example, distributed learning can be used on a network of phones, each of which have a local machine learning model used for auto-complete suggestions. Instead of requiring each phone to send private data such as text messages or emails to a central location for centralized learning and storage, a machine learning model can be passed directly from phone to phone. The traveling model is updated directly from the local model, without touching the local data, and the local model may also learn from the traveling model. Differential privacy or other formal privacy measures may then be applied to ensure that the most recent learned update has not exposed the individual who contributed data, using techniques such as secure aggregation (Bonawitz et al. 2017).

Distributed learning addresses the problem of the leaky data lake, the problem of data and power concentration, and the problem of exclusive ownership of the trained machine learning model. However, distributed learning is still a hierarchical system, the aims of which are set at the top. It does not give its diverse users the ability to form their own aims, or to build coalitions with one another to solve learning

⁹ For the purposes of this section, imagine the metaphorical “data lake” not as a peaceful reservoir but as a toxic tailings pond prone to catastrophic rupture.

problems of mutual interest. The objective of the model and the goals of its learning are set by the person or entity that designs the architecture of the machine learning model and sets it traveling. In order to imagine a more pluralistic set of tasks, we must add the ability for distributed agents to set more than one learning goal simultaneously—namely, multitask federating learning (Caruana 1997).

Federated learning is characterized by data that remains local and by a model development process that is distributed. Rather than streaming raw data towards a central site for interpretation, data remains on the user's device and model development occurs either (depending on the extent of the federation) on that device, or in a central location based only on the formatted, anonymized, and already-minimized data collected from the user. Because the raw data remains with the user, the user is both less tied into and less dependent on the centralized model and thereby transfers less control to the system's developers (Kairouz et al. 2021). This process can, if implemented properly, allow representation with fewer risks of exploitation. However, while control over the workings and answers remain with the user, the problem to be solved is often determined centrally (Kairouz et al. 2021).

Multitask federated learning improves on this model by allowing each person not only to maintain access to their data and to choose what learning to allow but also to contribute to learning goals of their own or others' devising (Kairouz et al. 2021). In doing so, it affords the possibility of pluralist machine learning systems. Multitask federated learning on its own is not a "fix" to the issue of who, and whose knowledge, counts. It does nothing to address the generation of data or the social valuation of knowers. Nevertheless, for researchers interested in forming a critical technical practice by hybridizing feminist theory and machine learning, this model provides one way to address some of the pragmatic concerns around power that stymie efforts to imagine feminist ML premised on a more conventional, centralized structure of AI.

3. Which Knowledge?

3.1. Eliding Difference in Cyc and Soar

When Alison Adam analyzed Cyc, she found a knowledge system shaped by the perspectives of the middle-class, white, male engineers who built it. Cyc's knowledge representations were not entirely univocal: the system did include the capacity to represent multiple competing models of the world. However, this capacity was reserved for cases of conflict between scientific theories "judged to be of similar intellectual status," such as competing theories within economics or current scientific theories, and superseded theories still used for teaching, such as Newtonian physics (Adam 1998, 85).

What Cyc did not model was the existence of multiple, observer-relative perspectives of the same event or the interplay between such perspectives, as in

Sandra Harding’s (1992) strong objectivity. When Cyc stores multiple conflicting theories or models, at least one must be coded as mere “belief” rather than “knowledge” unless the engineers believe the domain itself to be “inexact,” as with economics (Adam 1998, 87). Typically, the lower status “beliefs” represent minority opinions (Adam 1998, 88). A person who disagreed with Cyc’s judgements, or whose commonsense beliefs about the world were framed in a different way, would have little foothold from which to contest it.

Soar, without an extensive knowledge corpus, homogenized its problem-solving methodology instead. It sought to derive general problem-solving principles from Newell and Simon’s studies of male college students and model them in an artificial system. Newell and Simon believed that the goal-directed motivation, individual approach, and biological “normality” the undergraduates displayed constituted preconditions for rational problem-solving (Adam 1998, 96). However, their study did not seek to study other forms of human problem-solving or model them within Soar. In what follows, we will bring Adam’s critiques of Cyc and Soar into the pluralistic present and propose contemporary models that embrace rather than elide multiplicity.

3.2. Pluralism in the Present

Machine learning systems in public life have two basic modes: universalizing and personalizing. Systems typically aim either to universalize, to distill statistical patterns that reflect what “most people” do, or to personalize, learning information about each individual in order to better accord with their preferences. Some systems, such as biometric systems that seek to identify the individual through purportedly universal criteria, can do both (van der Ploeg 2011). Machine learning’s universalizing mode is the one critiqued by Adam for representing the perspective of majoritarian or socially dominant groups as the universal or default perspective. The personalizing mode has been critiqued as leading to polarization and to the creation of partisan “echo chambers.”¹⁰ These basic models each encourage different relations to perspective-taking and to knowledge. The universalizing model encourages users to recognize and orient themselves around an outside perspective—but it is that of a generalized, idealized version of a socially dominant group. The personalized mode re-presents one’s own perspective, eliding the existence of difference and the possibility of the “world-travelling” or “role-taking” that underpins much feminist theorizing about the nature of social relations and politics (Lugones 1987; Weir 2013).

Search engines, for example, assume that most people want one thing. When they type “Feminist Philosophy Quarterly” into a search engine, they want to find the

¹⁰ As described in the work of Nguyen (2020), although the empirical studies of political polarization attribute different causes.

website of the journal *Feminist Philosophy Quarterly*.¹¹ The desires of the remaining people are multifold. Many want to find individual articles within *FPQ* or authors who frequent its pages; some want to find submission instructions; some want to find *Hypatia's* website; some want to find articles critiquing *Feminist Philosophy Quarterly*; still others are disappointed to find that the name they picked out for their next journal is already in use. But all see models based on the same underlying criteria of relevance.

The singular ontological viewpoint a universal ranking represents often leads to the gaze of the majority trumping the needs of any particular individual and the reproduction of injustices ignored by majoritarian gazes. Safiya Noble (2018) presents a host of examples of search's prioritization of white and male searchers over Black and female searchers: image searches for "business attire" that return only white men in suits; searches for "Black girls" that return only erotica. In addition to their bias, these results represent a single perspective on what typifies "business attire" or "Black girls." The search engine's "view from nowhere" turns out to be a view from the perspective of dominant social groups, as Adam predicted would be the case.

The route to abandoning the single perspective and its biased universalism often wends through personalization. In a personalized model of search, each searcher would be shown "business attire" considered to be appropriate to them—or to how the search platform sees them. But in both cases, the ideal is treated as optimizing results to the user's most immediate needs, be they the user's actual needs or the needs of a fictional default. In neither situation, then, is there an effort to make the "road not taken," or the contingent and situated nature of the results offered, visible. While we use search as an example, the dichotomy of universalized versus personalized data processing and interpretation is ubiquitous, and in many cases desirable. A medical diagnosis system that intentionally does not recommend the most likely condition would be rightly abandoned.¹² But when systems bound and shape our senses of the social world and of each other, revealing the multiplicitous worlds and perspectives that are present is vital.

3.3. Machine Learning from Multiplicity: Critical Technical Practices

Intentionally revealing multiplicity and contingency in automated systems is an idea gaining momentum. Ochigame (2021) proposed "'divergent search,' which

¹¹ Perhaps the journal's open access treasures have hitherto been hidden from them, but not from you, at <https://ojs.lib.uwo.ca/index.php/fpq>.

¹² That said, medical recommendations made by automated diagnostic systems can benefit from displaying choices between values and choices guided by tolerance for risk that shape decisions made based on probabilities. For an example, see Birch et al. (2022).

seeks to facilitate exposure to divergent perspectives across linguistic and geographic barriers.” For example, Ochigame uses “divergent shuffle” to reorder search results so that the top ten listings include results from at least four regions, rather than nine out of ten results being from North America and Western Europe. Ochigame and Ye (2021) extend this work to build a divergent “search atlas.” Hancox-Li and Kumar also imagine a pluralistic machine learning in the context of feature choice when they say,

Given the uncertain relationships between those numbers [indicating importance of features] and the actual features in the data, visualizing [feature importance numbers] as though they are certain and have unambiguous importance values is misleading. For example, one can imagine an interface that includes multiple explanatory accounts of a model and helps users see the differences between them. In contrast, we currently have multiple, discrete explanation methods that each present their own seemingly authoritative accounts, hiding the uncertainty that is inherent in each of them. (Hancox-Li and Kumar 2021, 823–24)

Building machine learning systems with multiplicity renders visible different ranges of possibility and the perspectives they represent. Extending Ochigame & Ye’s and Hancox-Li & Kumar’s work, we propose automating these processes across a broader range of machine learning in public life, showcasing a multiplicity of perspectives and their contextual adaptation, using multitask learning, ensemble learning, and other multi-model learning methods.

The aim of showcasing multiple divergent perspectives can be accomplished using different degrees of “ontological” difference in the machine learning model’s internal representations. Consider, for example, different ways of implementing Ochigame’s divergent shuffle. The method closest to the status quo would be to maintain the same ranking of all papers, using the same criteria of relevance and the same learning task, but then to choose from that ranking papers to display based on additional optimization criteria. In the example Ochigame (2021) describes, a current search for scientific papers on “climate change” returns mostly papers from North America and Western Europe (NA) and one from Latin America (LA). Assume that the best paper from each region is labeled 1, the second best 2, and so on. Thus the current top ten search results are, in order, NA1, NA2, NA3, NA4, NA5, LA1, NA6, NA7, NA8, NA9. Divergent shuffle could draw from the same overall ranking of papers but instead show the searcher the following “shuffled” list, including papers from Africa (AF), Asia (AS), and Eastern Europe (EE): LA1, EE1, AF1, AS1, NA1, AS2, EE2, AF2, NA2, LA2. In order to implement this strategy, no additional machine learning techniques

are needed—it is only necessary to add an additional constraint on the results shown to the searcher.

A second step away from the status quo would be to train the machine learning model itself to optimize for multiple goals. Multitask learning, for example, allows different optimization tasks but maintains a shared internal representation (Caruana 1997). The different tasks in a pluralistic search would be providing the most “relevant” links to different people based on different criteria of relevance. Ranked lists could then be “shuffled” together, as in Ochigame’s divergent shuffle, so that more than one perspective is visible within one search.

The previous two methods rely on shared internal representations, data labels, and a shared (implicit or explicit) ontology. In order to present a more pluralistic pluralism, however, it may be necessary to allow different models that each rely on different data, data classified in different ways, or different learning methods. Federated learning, described in section 2.3, can be used to create multiple, heterogeneous models that can be synthesized into a global one (Diao, Ding, and Tarokh 2021). But a global model is not always necessary.

Ensemble learning is a suite of techniques that uses multiple, distinct machine learning models to perform the same task, then aggregates their results (Dietterich 2000). Ensemble learning often delivers better results than one model alone could, especially for complex decision landscapes in which a single model is likely to get stuck in a local maximum (Kairouz et al. 2021). Nina Grgić-Hlača and her coauthors (Grgić-Hlača et al. 2017), however, propose forgoing the aggregation step common to ensemble learning and instead choosing randomly between the results of the models for each token-decision instance. This preserves a diversity of results and a diversity of methods, albeit at the potential cost of giving up some “performance” on any single task.

Machine learning need not be a “one-world world” (Law 2015). Ensemble methods are mature and well developed. They can be purposed to serve pluralism rather than to increase performance on a single task.

4. What Knowledge?

4.1. Autonomy and Interdependence in Cyc and Soar

In addition to asking who knows and what they know, Adam questions the autonomy of knowers themselves. Our epistemic reliance on others begins in childhood. Years of dependency on others creates our “second” personhood, our self that is constituted in relationship to others (Baier 1985). Even as adults, much of our knowledge is from testimony or is deeply relational. As members of teams and partnerships, we rely on collective knowledge to perform tasks none of us could do individually. Thus the perceived tradeoff between interdependence and autonomy is

often illusory: interdependence expands our capacities whether we realize it or not (Code 1991, 79)

Drawing on Annette Baier (1985) and Lorraine Code (1991), Adam argues that as human persons are “second persons” whose knowledge is relational, they are not (nor should they be) fully autonomous epistemic agents. Given this, artificial systems modeled on human intelligence should neither assume that humans are fully autonomous nor strive for autonomous self-reliance themselves. Their goals should not include complying with the normative ideal of autonomy.

The symbolic systems that Adam critiqued strove for autonomy. Soar was modeled after humans who solved problems entirely on their own in the artificially isolated test setting of the laboratory (Adam 1998, 97). The undergraduates studied were not allowed to rely on connected knowing (Belenky et al. 1986), and so neither did Soar.

Cyc and Soar’s self-sufficiency are also a poor basis for the attribution of responsibility. As Adam argues, many disasters lack a single author (Adam 1998, 97–98). A system has failed when an oil spill destroys a coastline. Rather than resolving reasonable disagreement about who is to blame, Adam argues that members of the system should take collective responsibility.

4.2. Autonomy in the Present

The widespread deployment of AI systems has brought concerns about autonomy to a wider audience. In addition to promoting the normative ideal of autonomy highlighted by Adam, the political economy of automation has long incentivized the development of these technologies because of their promises to strip human discretion and decision-making from processes (Wajcman 2017; Feenberg 1991). That these promises are false—that these technologies are “humans all the way down” (Keyes 2018; Neyland 2019; Keyes 2020; Muller et al. 2021)—does not change the impact that both promises and technologies had and have.

Autonomy-oriented critiques of algorithmic systems usually examine one or both of two domains: the cultural imaginary of algorithms and what we might call their everyday life. Inquiries into cultural imaginaries are inquiries into narratives that “describe attainable futures and prescribe the images of futures that should be attained” (Felt et al. 2016, 754). Such narratives “condition not only the perception of technology within the public but also ‘the professional culture of those who have produced the technical innovations and helped their development’” (Natale and Ballatore, 2020, 6; quoting Ortoleva, 2009, 2). Such cultural imaginaries play a strong role in how we engage with and interpret events and each other (Babbitt 2018; Lindemann Nelson 2001).

Feminist scholars highlight the cultural imaginaries of AI to emphasize the ways in which algorithmic systems, regardless of their actual, material state or level

of integration may constrain our autonomy by constraining our range of imagined possibility. In a culture in which algorithms are portrayed as better than humans at decision-making or evaluation—seen as capable of inferring truths undetectable to humans—people are reluctant to challenge them. An algorithmic decision is less likely to be challenged than a human decision—not because it cannot be but because the algorithm is afforded a particular epistemic authority (Beer 2017). This is particularly worrisome with the increasing integration of AI into the production of cultural imaginaries and into the generation of “truths” around identity, legitimacy, or importance (Keyes, Hitzig, and Blell 2021).

Theorists also critique the everyday lives of algorithms: the day-to-day practices of their development and use (Neyland 2019). This work highlights both the increasing nonautonomous deployments of algorithmic systems, particularly in workplaces (Watkins 2021; Stark and Pais 2020), and the structuring of these systems in such a way as to exclude human agency and knowledge from informing their decision-making (Rubel, Castro, and Pham 2020). Designing AI systems to be central to decision-making reduces the autonomy of those interacting with such systems. Beyond the question of imaginaries, much algorithmic development and use still follows the pattern highlighted and critiqued by Adam—that of a monolithic system that simply provides “the answer,” without possibilities for user interrogation or involvement.

4.3. Relational Knowledge in the Loop: Critical Technical Practices

In contrast to monolithic systems, Adam sketched a vision of an artificial decision aid that leaves the decision open, a legal expert system flexible enough to advise by analogy. This vision of a process in which AI advises and humans decide can be seen in a model more broadly applicable to AI: the human-in-the-loop. Human-in-the-loop is a term used for a variety of human-machine collaborative decision-making: machine learning that relies on humans to label unlabeled data, identify edge cases that stumped the learning algorithm, and otherwise facilitate learning, but also automated decision-making that pauses at critical moments to allow the human to decide. (Looney and Tacker 1990; Falcone and Castelfranchi 2001; Enarsson, Enqvist, and Naarttijärvi 2022)

Being a human-in-the-loop is itself educational. Those who see the capacities and limitations of an algorithmic system learn to place appropriate trust in the system, learning when to rely on the system and when to rely on their own capacities (Abdel-Karim et al. 2020). Such systems therefore have the potential to undercut the cultural mythology that elevates the capacities of AI above human capacities and to increase human confidence in disputing algorithmic systems.

Indeed, being a human-in-the-loop often lowers trust in the system, a fact that is sometimes seen as a reason to shield humans from the loop (Honeycutt, Nourani,

and Ragan 2020). This reasoning assumes that trust is an unquestioned good, an assumption that many feminist and political philosophers would question. Warranted trust is certainly beneficial—but trust that is unwarranted can lead to overreliance on the other party and to harm when that trust is violated. Further, active distrust is often seen as a foundational part of rendering systems accountable, be they social or sociotechnical. An attitude of distrust—an attitude in which we approach situations with a degree of suspicion—reveals flaws and encourages the desire to improve (Rosanvallon 2008).

Rather than simply seeking to increase user trust in automated systems, a better goal for system designers would be to allow users to appropriately calibrate trust to the capacities and limitations of the system. If users begin with an unrealistically high trust in the system, its capacities and objectivity, observing the system’s inevitable stumbles will decrease their trust. But this is epistemically appropriate. Seeing the brittle edges of automated knowledge allows humans-in-the-loop to increase their comparative trust in themselves.

What might critical technical practices around trust, autonomy and second-persons look like, then? We would argue that a vital part of demystification is exposure to and involvement with feminist AI and its potential to render visible the mechanisms of algorithmic systems. We point to the ongoing work to create feminist makerspaces and hackathons—sites of deliberate, collaborative making and learning about technologies (Fox, Silva, and Rosner 2018; Houston et al. 2016). These environments are hardly perfect; they have their own dynamics of power around gender, race, and class. But they constitute a starting point for moving beyond monolithic imaginaries of AI.

Similar proposals are made by D’Ignazio and Klein (2020), who highlight feminist data mapping projects in their work on data feminism. These activities are vital, but still leave “the algorithm” itself unquestioned. We urge practitioners to go beyond mapping alone and instead build spaces for the creation and deployment of models. Such spaces offer the possibility of deep experience with the fragility and multiplicity of algorithmic systems, and so they offer an alternative vision of the world—one in which epistemic deference to AI is weaker and trust is given when warranted.

In addition to developing warranted trust, critical technical practices can also respond to—or preclude—trust’s violation. Leigh Star’s famous description of infrastructure as “invisible until breakdown” (Star and Ruhleder 1996, 113) carries with it a corollary: infrastructure is visible (and seemingly not “infrastructure” at all) to those inside the practices that make the infrastructure function.

Louise Amoore argues that a certain amount of unknowability-of-outcomes is inevitable—not just within AI, but in interaction and relation more generally (Amoore 2020). Correspondingly, there will always be unforeseen violations of trust. Our

response should not be to mandate full transparency (which is, as she argues, impossible) but instead to develop a “cloud ethics”: an ethicopolitical approach that includes denaturalizing the choices that have led to a particular algorithm, problem, or solution being the one actively developed by “dwell[ing] for some time with the aperture of the algorithm, the point where the vast multiplicity of parameters and hidden layers becomes reduced and condensed to the thing of interest” (Amoore 2020, 162). Purposefully embedding human agents in an algorithmic system gives them an inherently partial and reactive epistemic access to its functioning; people respond to systems as much as the other way around. But this embedding carries the potential to make those apertures, for the people embedded, visible; to enable precisely the kind of dwelling for which Amoore advocates, and through that, to enable new ways to preclude or respond to algorithmic harms.¹³

5. Conclusion

Despite changes in the systems and technical capacities of AI and machine learning in the last thirty years, feminist philosophy’s critiques remain relevant. A world in which algorithmic knowledge is pluralistic and localized (when appropriate), in which humans trust in and question algorithmic systems to the degree warranted, and in which neither humans nor machines are viewed as autonomous epistemic agents has been imaginable for a long time. And this history in itself can be a source of hope. Like Adam (1998, 181), we are “telling one more version of an old story,” and with the same aim: to show that although neither our projects nor our problems are new, by “continuing to build on the practical projects just begun, and through women’s refusal to give up ground made in relation to technology, we gain a glimpse, however small, of how things could be different” (181).

Acknowledgments

The authors would like to thank Claire and Margaret Hopkins, Margaret Wander, and Shannon Hackett for their support; Liam Kofi Bright, Roger Creel, Kathleen Nichols, and Olivia Ordoñez for helpful comments; the attendees and organizers of the “Feminism, Social Justice, and AI” workshop for their thoughtful engagement with the early versions of this paper; and, of course, Alison Adam.

References

Abdel-Karim, Benjamin M., Nicolas Pfeuffer, Gernot Rohde, and Oliver Hinz. 2020. “How and What Can Humans Learn from Being in the Loop?” *KI-Künstliche*

¹³ Experiments in enabling precisely this kind of ethicopolitics are under way in sociotechnical infrastructures of science; see Inman and Ribes (2018).

- Intelligenz* 34, no. 2 (June): 199–207. <https://doi.org/10.1007/s13218-020-00638-x>.
- Adam, Alison. 1998. *Artificial Knowing: Gender and the Thinking Machine*. London: Routledge.
- Agre, Phillip E. 2014. "Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI." In *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, edited by Geoffrey C. Bowker, Susan Leigh Star, William Turner, and Les Gasser, ch. 6. New York: Psychology Press.
- Allhutter, Doris. 2019. "Of 'Working Ontologists' and 'High-Quality Human Components': The Politics of Semantic Infrastructures." In *digitalSTS: A Field Guide for Science & Technology Studies*, edited by Janet Vertesi and David Ribes, 326–48. Princeton, NJ: Princeton University Press. <https://digitalsts.net/essays/of-working-ontologists-and-high-quality-human-components/>.
- Amoore, Louise. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.
- . 2022. "Machine Learning Political Orders." *Review of International Studies*. Published ahead of print, February 15, 2022. <https://doi.org/10.1017/S0260210522000031>.
- Babbitt, Susan E. 2018. *Impossible Dreams: Rationality, Integrity, and Moral Imagination*. New York: Routledge.
- Baier, Annette. 1985. *Postures of the Mind: Essays on Mind and Morals*. Minneapolis: University of Minnesota Press.
- Barth-Jones, Daniel. 2012. "The 'Re-identification' of Governor William Weld's Medical Information: A Critical Re-examination of Health Data Identification Risks and Privacy Protections, Then and Now." Unpublished paper posted on SSRN. Revised September 3, 2015. <https://doi.org/10.2139/ssrn.2076397>.
- Beer, David. 2017. "The Social Power of Algorithms." *Information, Communication & Society* 20 (1): 1–13.
- Belenky, Mary Field, Blythe McVicker Clinchy, Nancy Rule Goldberger, and Jill Mattuck Tarule. 1986. *Women's Ways of Knowing: The Development of Self, Voice, and Mind*. New York: Basic Books.
- Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.
- Birch, Jonathan, Kathleen A. Creel, Abhinav K. Jha, and Anya Plutynski. 2022. "Clinical Decisions Using AI Must Consider Patient Values." *Nature Medicine* 28, no. 2 (February): 229–32. <https://doi.org/10.1038/s41591-021-01624-y>.
- Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. 2017. "Practical Secure Aggregation for Privacy-Preserving Machine Learning." In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and*

- Communications Security*, 1175–91. New York: Association for Computing Machinery. <https://doi.org/10.1145/3133956.3133982>.
- Browne, Simone. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press.
- Bucher, Taina. 2018. *If . . . Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Caruana, Rich. 1997. "Multitask Learning." *Machine Learning* 28, no. 1 (July): 41–75. <https://doi.org/10.1023/A:1007379606734>,
- Christen, Kimberly A. 2012. "Does Information Really Want to Be Free? Indigenous Knowledge Systems and the Question of Openness." *International Journal of Communication* 6:2870–93. <https://ijoc.org/index.php/ijoc/article/view/1618/828>.
- Cifor, Marika, Patricia Garcia, T. L. Cowan, Jasmine Rault, Tonia Sutherland, Anita Say Chan, Jennifer Rode, Anna Lauren Hoffmann, Niloufar Salehi, and Lisa Nakamura. 2019. "Feminist Data Manifest-No." <https://www.manifestno.com/>.
- Code, Lorraine. 1991. *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca, NY: Cornell University Press.
- . 2012. "Taking Subjectivity into Account." In *Education, Culture and Epistemological Diversity: Mapping a Disputed Terrain*, edited by Claudia W. Ruitenberg and D. C. Phillips, 85–100. Dordrecht: Springer.
- Diao, Enmao, Jie Ding, and Vahid Tarokh. 2021. "HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients." Paper presented at International Conference on Learning Representations, virtual event, Vienna, May 3–7, 2021.
- Dietterich, Thomas G. 2000. "Ensemble Methods in Machine Learning." In *Multiple Classifier Systems: First International Workshop, MCS 2000, Cagliari, Italy, June 21–23, 2000, Proceedings*, edited by Josef Kittler and Fabio Roli, 1–15. Berlin: Springer.
- D'Ignazio, Catherine. and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, MA: MIT Press.
- Dreyfus, Hubert L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.
- Dwork, Cynthia. 2008. "Differential Privacy: A Survey of Results." In *Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 2008, Proceedings*, edited by Maninda Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li, 1–19. Berlin: Springer.
- Enarsson, Therese, Lena Enqvist, and Markus Naarttijärvi. 2022. "Approaching the Human in the Loop—Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts." *Information & Communications Technology Law* 31 (1): 123–53. <https://doi.org/10.1080/13600834.2021.1958860>.

- Falcone, Rino, and Cristiano Castelfranchi. 2001. "The Human in the Loop of a Delegated Agent: The Theory of Adjustable Social Autonomy." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 31, no. 5 (September): 406–18. <https://doi.org/10.1109/3468.952715>.
- Feenberg, Andrew. 1991. *Critical Theory of Technology*. New York: Oxford University Press.
- Felt, Ulrike, Judith Igelsböck, Andrea Schikowitz, and Thomas Völker. 2016. "Transdisciplinary sustainability research in practice: between imaginaries of collective experimentation and entrenched academic value orders." *Science, Technology, & Human Values* 41, no. 4 (July): 732–61. <https://doi.org/10.1177/0162243915626989>.
- Ford, Heather, and Judy Wajcman. 2017. "'Anyone Can Edit', Not Everyone Does: Wikipedia's Infrastructure and the Gender Gap." *Social Studies of Science* 47, no. 4 (August): 511–27. <https://doi.org/10.1177/0306312717692172>.
- Fox, Sarah E., Rafael M. L. Silva, and Daniela K. Rosner. 2018. "Beyond the Prototype: Maintenance, Collective Responsibility, and Public IoT." In *DIS '18: Proceedings of the 2018 Designing Interactive Systems Conference*, 21–32. New York: Association for Computing Machinery. <https://doi.org/10.1145/3196709.3196710>.
- Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston, MA: Houghton Mifflin Harcourt.
- Grgić-Hlača, Nina, Muhammad Bilal Zafar, Krishna P. Gummadi, and Adrian Weller. 2017. "On Fairness, Diversity and Randomness in Algorithmic Decision Making." arXiv preprint, arXiv:1706.10208 [stat.ML]. <https://doi.org/10.48550/arXiv.1706.10208>.
- Hancox-Li, Leif, and I. Elizabeth Kumar. 2021. "Epistemic Values in Feature Importance Methods: Lessons from Feminist Epistemology." In *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 817–26. New York: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445943>.
- Harding, Sandra. 1992. "Rethinking Standpoint Epistemology: What is 'Strong objectivity'?" *Centennial Review* 36, no. 3 (Fall): 437–70.
- Honeycutt, Donald R., Mahsan Nourani, and Eric D. Ragan. 2020. "Soliciting Human-in-the-Loop User Feedback for Interactive Machine Learning Reduces User Trust and Impressions of Model Accuracy." In *Proceedings of the Eighth AAAI Conference on Human Computation and Crowdsourcing*, edited by Lora Aroyo and Elena Simperl, 63–72. Palo Alto, CA: Association for the Advancement of Artificial Intelligence. <https://ojs.aaai.org/index.php/HCOMP/article/view/7464>.
- Houston, Lara, Steven J. Jackson, Daniela K. Rosner, Syed Ishtiaque Ahmed, Meg Young, and Laewoo Kang. 2016. "Values in Repair." In *CHI '16: Proceedings of*

- the 2016 CHI Conference on Human Factors in Computing Systems*, 1403–14. New York: Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858470>.
- Hutson, Jevan A., Jessie G. Taft, Solon Barocas, and Karen Levy. 2018. “Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms.” *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW (November): article no. 73, 1–18. <https://doi.org/10.1145/3274342>.
- Inman, Sarah, and David Ribes. 2018. “Data Streams, Data Seams: Toward a Seamless Representation of Data Interoperability.” In *Proceedings of DRS 2018 International Conference, Vol. 1: Design as a Catalyst for Change*, edited by Cristiano Storni, Keelin Leahy, Muireann McMahon, Peter Lloyd and Erik Bohemia, 35–49. London: Design Research Society. <https://doi.org/10.21606/drs.2018.359>.
- Introna, Lucas D., and Helen Nissenbaum. 2000. “Shaping the Web: Why the Politics of Search Engines Matters.” *Information Society* 16 (3): 169–85. <https://doi.org/10.1080/01972240050133634>.
- Irani, Lilly C., and M. Six Silberman. 2013. “Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk.” In *CHI '13, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–20. New York: Association for Computing Machinery. <https://doi.org/10.1145/2470654.2470742>.
- Kairouz, Peter, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz et al. 2021. “Advances and Open Problems in Federated Learning.” *Foundations and Trends® in Machine Learning* 14 (1–2): 1–210. <https://doi.org/10.1561/22000000083>.
- Keys, Os. 2018. “The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition.” *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW (November): article no. 88, 1–22. <https://doi.org/10.1145/3274357>.
- . 2020. “Automating Autism: Disability, Discourse, and Artificial Intelligence.” *Journal of Sociotechnical Critique* 1 (1): article 8, 1–31. <https://doi.org/10.25779/89bj-j396>.
- Keys, Os, Zoë Hitzig, and Mwenza Blell. 2021. “Truth from the Machine: Artificial Intelligence and the Materialization of Identity.” *Interdisciplinary Science Reviews* 46 (1–2): 158–75. <https://doi.org/10.1080/03080188.2020.1840224>.
- Knight, Will. 2016. “An AI with 30 Years’ Worth of Knowledge Finally Goes to Work.” *MIT Technology Review*, March 14, 2016. <https://www.technologyreview.com/2016/03/14/108873/an-ai-with-30-years-worth-of-knowledge-finally-goes-to-work>.

- Law, John. 2015. "What's Wrong with a One-World World?" *Distinktion: Scandinavian Journal of Social Theory* 16 (1): 126–39. <https://doi.org/10.1080/1600910X.2015.1020066>.
- Lenat, Douglas B., and R. V. Guha. 1990. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Reading, MA: Addison-Wesley.
- Lindemann Nelson, Hilde. 2001. *Damaged Identities, Narrative Repair*. Ithaca, NY: Cornell University Press.
- Looney, Carl G., and Edgar C. Tacker. 1990. "Human-in-the-Loop Control with Majority Vote Neural Networks." In *1990 IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings*, 224–26. Los Angeles: IEEE. <https://doi.org/10.1109/ICSMC.1990.142097>.
- Lugones, María. 1987. "Playfulness, 'World'-Travelling, and Loving Perception." *Hypatia* 2, no. 2 (Summer): 3–19. <https://doi.org/10.1111/j.1527-2001.1987.tb01062.x>.
- Mager, Astrid. 2014. "Defining Algorithmic Ideology: Using Ideology Critique to Scrutinize Corporate Search Engines." *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 12 (1): 28–39. <https://doi.org/10.31269/triplec.v12i1.439>.
- Merler, Michele, Nalini Ratha, Rogerio S. Feris, and John R. Smith. 2019. "Diversity in Faces." arXiv preprint, arXiv:1901.10436 [cs.CV]. <https://doi.org/10.48550/arXiv.1901.10436>.
- Muller, Michael, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma et al. 2021. "Designing Ground Truth and the Social Life of Labels." In *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, article no. 94, 1–16. New York: Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445402>.
- Nagel, Thomas. 1989. *The View from Nowhere*. New York: Oxford University Press.
- Natale, Simone, and Andrea Ballatore. 2020. "Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence." *Convergence* 26, no. 1 (February): 3–18. <https://doi.org/10.1177/1354856517715164>.
- Neyland, Daniel. 2019. *The Everyday Life of an Algorithm*. Cham: Springer Nature.
- Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17, no. 2 (June): 141–61. <https://doi.org/10.1017/epi.2018.32>.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression*. New York: NYU Press.
- Ochigame, Rodrigo. 2021. "A Preview of Divergent Search." In *ATMI: Affecting Technologies, Machining Intelligences*, edited by Dalida María Benfield, Bruno Moreschi, Gabriel Pereira, and Katherine Ye. Center for Arts, Design and Social Research. <https://book.affecting-technologies.org/preview/>.

- Ochigame, Rodrigo, and Katherine Ye. 2021. Search Atlas: Visualizing divergent search results across geopolitical borders. In *DIS '21: Designing Interactive Systems Conference 2021*, edited by Wendy Ju, Lora Oehlberg, Sean Follmer, Sarah Fox, and Stacey Kuznetsov, 1970–83. New York: Association for Computing Machinery. <https://doi.org/10.1145/3461778.3462032>.
- Ortoleva, Peppino. 2009. "Modern Mythologies, the Media and the Social Presence of Technology." *Observatorio (OBS) Journal* 3 (1): 1–12. <https://obs.obercom.pt/index.php/obs/article/view/163>.
- Rincón, Cami, Os Keyes, and Corinne Cath. 2021. "Speaking from Experience: Trans/Non-Binary Requirements for Voice-Activated AI." *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1 (April): article no. 132, 1–27. <https://doi.org/10.1145/3449206>.
- Rogaway, Phillip. 2015. "The Moral Character of Cryptographic Work." Unpublished manuscript.
- Rosanvallon, Pierre. 2008. *Counter-Democracy: Politics in an Age of Distrust*. Translated by Arthur Goldhammer. Cambridge: Cambridge University Press.
- Rubel, Alan, Clinton Castro, and Adam Pham. 2020. "Algorithms, Agency, and Respect for Persons." *Social Theory and Practice* 46, no. 3 (July): 547–72. <https://doi.org/10.5840/soctheorpract202062497>.
- Sadowski, Jathan. 2019. "When Data is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6, no. 1 (January): 1–12. <https://doi.org/10.1177/2053951718820549>.
- Sherron, Catherine. 2000. "Constructing Common Sense." In *Women, Work and Computerization: Charting a Course to the Future*, edited by Ellen Balka and Richard Smith, 111–18. Boston: Springer.
- Soon, Valerie. 2021. "Social Structural Explanation." *Philosophy Compass* 16, no. 10 (October): e12782. <https://doi.org/10.1111/phc3.12782>.
- Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7, no. 1 (March): 111–34. <https://doi.org/10.1287/isre.7.1.111>.
- Stark, David, and Ivana Pais. 2020. "Algorithmic Management in the Platform Economy." *Sociologica* 14 (3): 47–72. <https://doi.org/10.6092/issn.1971-8853/12221>.
- Stark, Luke. 2018. "Facial Recognition, Emotion and Race in Animated Social Media." *First Monday* 23, no. 9 (September). <https://doi.org/10.5210/fm.v23i9.9406>.
- Stevens, Nikki, and Os Keyes. 2021. "Seeing Infrastructure: Race, Facial Recognition and the Politics of Data." *Cultural Studies* 35 (4–5): 833–53. <https://doi.org/10.1080/09502386.2021.1895252>.
- Sweeney, Latanya. 2000. "Simple Demographics Often Identify People Uniquely." Data Privacy Working Paper, No. 3: Carnegie Mellon University.

- van der Ploeg, Irma. 2011. "Normative Assumptions in Biometrics: On Bodily Differences and Automated Classifications." In *Innovating Government: Normative, Policy and Technological Dimensions of Modern Government*, edited by Simone van der Hof and Marga M. Groothuis, 29–40. The Hague: T. M. C. Asser Press.
- Vrandečić, Denny, and Markus Krötzsch. 2014. "Wikidata: A Free Collaborative Knowledgebase." *Communications of the ACM* 57, no. 10 (October): 78–85. <https://doi.org/10.1145/2629489>.
- Wajcman, Judy. 2017. "Automation: Is It Really Different This Time?" *British Journal of Sociology* 68, no. 1 (March): 119–27. <https://doi.org/10.1111/1468-4446.12239>.
- Watkins, Elizabeth Anne. 2021. "The Polysemia of Recognition: Facial Recognition in Algorithmic Management." PhD diss., Columbia University.
- Weir, Allison. 2013. *Identities and Freedom: Feminist Theory between Power and Connection*. Oxford: Oxford University Press.

OS KEYES is a PhD candidate at the University of Washington, where they study AI and biomedicine at the intersect of trans studies, philosophy, science and technology studies, and critical disability studies. They are an inaugural recipient of the Ada Lovelace Fellowship.

KATHLEEN CREEL is an assistant professor at Northeastern University, cross-appointed between the Department of Philosophy and Religion and Khoury College of Computer Sciences. She was previously an Embedded EthiCS fellow at Stanford University. Her research interests include philosophy of science, ethics and epistemology of machine learning, and early modern philosophy.